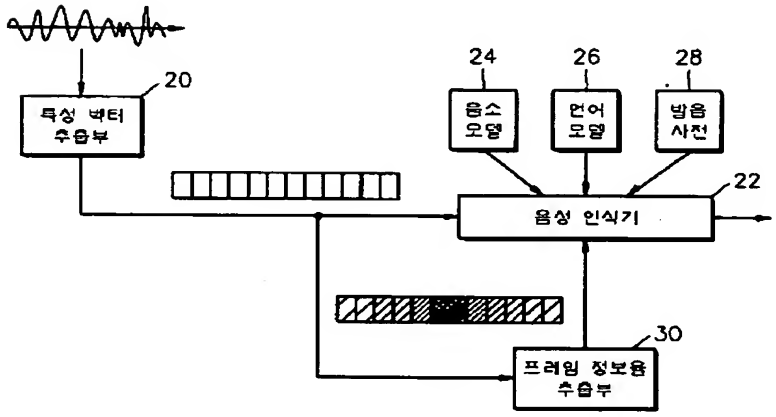
KOREAN PATENT ABSTRACTS(KR)

Document Code:A
(11) Publication No.1020010088054 (43) Publication.Date. 20010926
(21) Application No.1020000012052 (22) Application Date. 20000310
(51) IPC Code:
G10L 15/14

(71) Applicant:
SAMSUNG ELECTRONICS CO., LTD.
(72) Inventor:
CHOI, IN JEONG
KIM, DO YEONG
(30) Priority:

(54) Title of Invention
SPEECH RECOGNITION DEVICE AND METHOD USING WEIGHING VALUE PER STATE

Representative drawing



(57) Abstract:

PURPOSE: A speech recognition device and a method using weighing values per state are provided to be capable of remarkably reducing error recognition rate of speech by applying a context dependent state weighing value.

CONSTITUTION: A characteristic vector extractor(20) receives speech signals of a speaker, extracts the characteristic vector of the speech signal at a fixed frame rate, and adds same weighing values to the characteristic vectors pre the extracted frame. A frame information rate extractor(30) applies weighing value per hidden makov states to respective frames through a multi layer percept theory learned by a divisional method from a learning database, and

generates a characteristic vector with a context dependent information rate for respective frames. A speech recognizer(22) recognizes the characteristic vector with a context dependent information rate using a hidden makov model and a learning model.

(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(51) Int. Cl.[7]
G10L 15/14

(45) 공고일자 2002년03월13일
(11) 등록번호 10-0327486
(24) 등록일자 2002년02월22일

(21) 출원번호 10-2000-0012052
(22) 출원일자 2000년03월10일
(65) 공개번호 특2001-0088054
(43) 공개일자 2001년09월26일

(73) 특허권자 삼성전자 주식회사
경기 수원시 팔달구 매탄3동 416
(72) 발명자 최인정
경기도수원시권선구권선동삼천리권선2차아파트103동507호
김도영
경기도수원시권선구권선동권선현대아파트204동902호
(74) 대리인 이영필, 조혁근, 이해영

심사관 : 남인호

(54) 스테이트별 가중치를 적용한 음성 인식 장치 및 방법

요약

스테이트별 가중치를 적용한 음성 인식 장치 및 방법이 공개된다. 음향 문맥에 상응하여 특성 벡터마다 히든 마코프 스테이트별로 특성 벡터의 중요성을 판단하여 화자의 음성을 인식하는 본 발명에 따른 음성 인식 장치는 화자의 음성 신호를 받아들이고, 고정 프레임 율로 음성 신호의 특성 벡터를 추출하고, 추출된 프레임별 특성 벡터에 동일한 가중치를 부여하는 특성 추출부, 학습 데이터 베이스로부터 분별적 방법에 의해 학습된 다층 퍼셉트론에 의해 프레임마다 히든 마코프 스테이트별 가중치를 적용하여, 프레임별로 문맥 의존적인 정보율을 갖는 특성 벡터를 발생하는 프레임 정보율 추출부 및 히든 마코프 모델과 학습 모델들을 이용하여 학습된 다층 퍼셉트론으로부터 추정된 문맥 의존적인 스테이트별 가중치를 갖는 특성 벡터를 인식하는 음성 인식부를 포함하는 것을 특징으로 하고, 주변 음성의 변이 정도에 따라 스테이트별 가중치를 달리 부여하므로, 스테이트별로 정보적 중요도를 반영하여 음성을 인식하므로 음성의 오인식률을 크게 줄일 수 있다.

대표도

도1

명세서

도면의 간단한 설명

도 1은 본 발명에 따른 문맥 의존적 스테이트별 가중치를 적용한 음성 인식 장치를 개략적으로 나타내는 블록도이다.

도 2(a)~(e)는 도 1에 도시된 장치의 각 부의 입/출력을 나타내는 도면이다.

도 3은 본 발명에 따른 음성 인식 방법을 나타내는 플로우 차트이다.

도 4는 다층 퍼셉트론을 이용한 본 발명에 따른 문맥 의존적인 스테이트별 가중치의 추정 장치를 나타내는 블록도이다.

도 5는 다층 퍼셉트론을 이용한 본 발명에 따른 문맥 의존적인 스테이트별 가중치의 추정 방법을 나타내는 플로우 차트이다.

도 6(a)~(d)는 본 발명에서와 같이 문맥 의존적인 스테이트별 가중치가 적용한 경우와 적용되지 않은 경우의 음성 인식 결과를 나타내는 도면이다.

도 7은 고립단어 태스크와 연속음성 태스크에 대해 종래 인식기와 본 발명간의 음성 인식 결과를 나타내는 도면이다.

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 음성 인식 장치에 관한 것으로, 특히, 문맥 의존적인 스테이트별 가중치를 적용한 음성 인식 장치 및 방법과 문맥 의존적인 스테이트별 가중치의 추정 장치 및 방법에 관한 것이다.

일반적인 음성 인식기는 각 프레임의 가중치가 동일하다고, 즉, 각 음성 구간이 인식에 미치는 영향은 같다는 가정하에 음성을 인식한다. 또한, 통계적인 방법에 근거하여 학습과 인식을 수행하기 때문에 상대적으로 길고 안정된 음성 구간들에 의해 모델이 학습되며, 그에 따라 인식 결과가 결정된다. 그러나, 실제로는 음성의 동적 특성들이 인식에 결정적인 역할을 한다고 알려져 있으며, 기존의 인식기들은 이러한 특성을 반영하는 데 많은 한계를 가지고 있다. 종래에, 각 음성 프레임별로 정보적 중요도 즉, 가중치를 반영하기 위하여 다음과 같은 두 가지 방법들이 제안되어 왔다.

첫번째 방법은 가변 프레임율 분석 방법이다. 가변 프레임율 분석 방법은 일정 간격으로 추출된 관측열로부터 변화가 심한 구간의 특징 벡터들은 그대로 인식에 사용하고, 안정된 구간에서는 일부 특징 벡터들만을 선별하여 인식에 이용하는 방법이다. 따라서, 상대적으로 변화가 많은 구간에서는 많은 특징 벡터들이 사용되어 중요시되고, 반대로 길고 안정된 구간에서는 적은 수의 특징 벡터들만 사용되기 때문에 인식에 미치는 중요도가 떨어진다.

이 방법은 인식에 사용되는 프레임의 수가 줄어들기 때문에 인식 시간이 짧다는 장점이 있다. 하지만, 이 방법은 중요 프레임들의 선택을 위한 명확한 기준이 결여되어 있다는 것과 실제 각 프레임들이 인식 대상 패턴들에 미치는 상대적인 중요성이 더 결정적인 요소임이 고려되고 있지 않다.

두 번째 방법은 기존의 음성 인식기에 스테이트별 가중치를 부여하는 방법이다. 즉, 모델간의 분별력을 높이려는 취지에서 히든 마코프 모델(HMM:Hidden Makov model)의 스테이별로 고정된 가중치를 부여한다. 그리고, 스테이트별 출력 확률의 로그값에 스테이트별 가중치가 곱해져서 얻어진 출력값이 인식에 이용된다. 스테이트별 가중치는 분별적 학습법을 포함한 여러 방법으로 추정되었으며, 실제로 이 방법을 적용하여 기존 인식기보다 더 나은 성능을 보여준다. 또한, 두 번째 방법은 각 모델 내의 스테이트간 가중치를 달리하기 때문에, 모델을 구성하는 특정 스테이트들이 다른 스테이트들보다 더 중요하다는 것이 표현된다. 그러나, 실제로 스테이트별 가중치는 고정된 값이 아니라 그 모델의 지속시간이나 변화 특성 등에 의해 변화되어야 한다. 따라서, 이처럼 스테이트별로 고정된 가중치가 부여된 경우 모델별 분별력이 개선되지 않는다는 단점이 있다.

스테이트별 가중치의 타당성을 증명한 것은 가변 정보율 모델을 이용한 음성 인식 방법에 대한 특허를 들 수 있다(국내 특허 95-18112). 이 발명에서는 음성 신호의 기본 구간들마다 그 구간의 음성 변이성에 따라 추출할 특징 벡터의 수를 달리하고자 가변 정보율 분석 모델을 제안하였다. 그러나, 구간별 특징 벡터의 수를 달리하는 접근 방법은 구현상 어려움이 있고, 구간별 중요성들이 크게 차이를 보이지 않기 때문에 실제 적용하기에 어려움이 있다.

발명이 이루고자하는 기술적 과제

본 발명이 이루고자 하는 제1기술적 과제는 문맥 의존적인 스테이트별 가중치를 적용한 음성 인식 장치를 제공하는 데 있다.

본 발명이 이루고자 하는 제2기술적 과제는 문맥 의존적인 스테이트별 가중치를 적용한 음성 인식 방법을 제공하는 데 있다.

본 발명이 이루고자 하는 제3기술적 과제는 문맥 의존적인 스테이트별 가중치의 학습 장치를 제공하는 데 있다.

본 발명이 이루고자 하는 제4기술적 과제는 문맥 의존적인 스테이트별 가중치의 학습 방법을 제공하는 데 있다.

발명의 구성 및 작용

상기 제1과제를 이루기 위해, 음향 문맥에 상응하여 특성 벡터마다 히든 마코프 스테이트별로 특성 벡터의 중요성을 판단하여 화자의 음성을 인식하는 본 발명에 따른 음성 인식 장치는 화자의 음성 신호를 받아들이고, 고정 프레임 율로 음성 신호의 특성 벡터를 추출하고, 추출된 프레임별 특성 벡터에 동일한 가중치를 부여하는 특성 추출부, 학습 데이터 베이스로부터 분별적 방법에 의해 학습된 다층 퍼셉트론에 의해 프레임마다 히든 마코프 스테이트별 가중치를 적용하여, 프레임별로 문맥 의존적인 정보율을 갖는 특성 벡터를 발생하는 프레임 정보율 추출부 및 히든 마코프 모델과 학습 모델들을 이용하여 학습된 다층 퍼셉트론으로부터 추정된 문맥 의존적인 스테이트별 가중치를 갖는 특성 벡터를 인식하는 음성 인식부를 포함하는 것이 바람직하다.

상기 제2과제를 이루기 위해, 음향 문맥에 따라 히든 마코프 스테이트별로 특성 벡터의 중요도를 판단하여 화자의 음성을 인식하는 본 발명에 따른 음성 인식 방법은 화자의 음성 신호로부터 고정 프레임 율로 특성 벡터를 추출하고, 추출된 프레임별 특성 벡터에 동일한 가중치를 부여하여 동일한 스테이트별 가중치를 갖는 프레임별 특성 벡터를 생성하는 (a)단계, 학습 데이터 베이스에 대해 분별적 방법에 의해 학습된 다층 퍼셉트론에 의해 (a)단계에서 추출된 프레임마다 히든 마코프 스테이트별 가중치를 부여하여 문맥 의존적인 스테이트별 가중치를 갖는 특성 벡터를 생성하는 (b)단계 및 히든 마코프 모델과 학습 모델들을 이용하여 학습된 다층 퍼셉트론으로부터 추정된 문맥 의존적인 스테이트별 가중치를 갖는 특성 벡터를 인식하는 (c)단계로 이루어지는 것이 바람직하다.

상기 제3과제를 이루기 위해, 학습을 통해 문맥 의존적인 스테이트별 가중치를 학습하는 본 발명에 따른 가중치 학습 장치는 현재 인식하고자 하는 학습 문장에 대한 특성 벡터와 스테이트별 가중치에 상응하여 복수개의 후보 가설들을 탐색하는 음성 인식부, 올바른 가설에 대한 확률 값과 음성 인식부에서 탐색된 후보 가설들에 대한 확률값 사이의 거리를 측정하여 오류 함수값을 계산하는 에러 정정부, 입력층, 출력층 및 중간층으로 구성되며, 현재 인식하고자 하는 학습 문장의 프레임을 중심으로 전/후 p개의 프레임을 받아들이고, 오류 함수값에 상응하여 입력층, 출력층 및 중간층간의 연결 가중치를 조정하는 다층 퍼셉트론 및 다층 퍼셉트론의 연결 가중치에 상응하여 현재 인식하고자 하는 학습 문장의 스테이트별 가중치를 추정하고, 추정된 스테이트별 가중치를 음성 인식부로 출력하는 가중치 추정부를 포함한다.

상기 제4과제를 이루기 위해, 다층 퍼셉트론을 이용하여 학습 문장의 문맥 의존적인 스테이트별 가중치를 학습하는 가중치 학습 방법은 문맥 의존적인 스테이트별 가중치들이 1에 가까운 값으로 초기화되도록 다층 퍼셉트론의 연결 가중치들을 0에 가까운 값으로 초기화하는 (a)단계, 현재 인식하고자 하는 학습 문장에 대한 특성 벡터와 스테이트별 가중치에 상응하여 학습 데이터 베이스로부터 복수개의 후보 가설들을 탐색하는 (b)단계, 올바른 가설에 대한 확률값과 (b)단계에서 탐색된 후보 가설들에 대한 확률값 사이의 거리를 측정하여 에러 함수값을 구하는 (c)단계, 에러 함수값에 상응하여 다층 퍼셉트론의 연결 가중치를 조정하는 (d)단계, 다수개의 학습 문장들에 대한 다층 퍼셉트론의 연결 가중치가 모두 조정되었는가를 판단하는 (e)단계 및 (e)단계에서 다수개의 학습 문장들에 대한 다층 퍼셉트론의 연결 가중치가 모두 조정되었다고 판단되면, 에러 함수값이 소정값 이하로 될 때까지 (b) 내지 (e)단계를 반복 진행하는 (f)단계로 이루어진다.

이하, 본 발명에 따른 문맥 의존적인 스테이트별 가중치를 적용한 음성 인식 장치 및 방법과 문맥 의존적인 스테이트별 가중치의 학습 장치 및 방법을 첨부한 도면들을 참조하여 다음과 같이 설명한다.

도 1은 본 발명에 따른 문맥 의존적 스테이트별 가중치를 적용한 음성 인식 장치를 개략적으로 나타내는 블록도이다. 본 발명에 따른 음성 인식 장치는 특성 벡터 추출부(20), 음성 인식부(22), 프레임 정보율 추출부(30)를 포함하여 구성된다. 설명의 편의를 위해, 도 1에는 음소 모델(24), 언어 모델(26), 발음 사전(28)등에 대한 학습 데이터 베이스를 함께 도시하였다.

도 1을 참조하여, 특성 벡터 추출부(20)는 화자의 음성 신호를 받아들인 후, 고정 프레임 율로 음성 신호의 특성 벡터를 추출하고, 추출된 프레임별 특성 벡터에 동일한 가중치를 부여한다.

프레임 정보율 추출부(30)는 학습 데이터 베이스로부터 분별적 방법에 의해 학습된 다층 퍼셉트론에 의해, 특성 벡터 추출부(20)에서 출력되는 동일한 가중치가 부여된 프레임별 특성 벡터에 문맥 의존적인 가중치를 부여한다. 따라서, 프레임 정보율 추출부(30)는 프레임별로 문맥 의존적인 정보율을 갖는 특성 벡터를 음성 인식부(22)로 출력한다. 학습 데이터 베이스

음성 인식부(22)는 히든 마코프 모델 방식을 이용하여, 음소 모델(24), 언어 모델(26) 및 발음사전(28) 등의 학습 모델들로부터 문맥 의존적인 스테이트별 가중치를 갖는 특성 벡터에 상응하는 모델을 탐색한다. 또한, 음성 인식부(22)는 탐색된 결과를 음성 인식 결과로서 출력 단자 OUT을 통해 출력한다.

도 2(a)~(e)는 도 1에 도시된 장치의 각 부의 입/출력을 나타내는 도면으로서, 도 2(a)는 화자의 음성 신호를 나타내고, 도 2(b)는 일정한 프레임별로 음성 신호의 특성 벡터를 추출하기 위한 윈도우를 나타내고, 도 2(c)는 고정된 스테이트별 가중치가 적용된 프레임별 정보율을 나타내고, 도 2(d)는 문맥 의존적인 스테이트별 가중치가 적용된 프레임별 정보율을 나타내고, 도 2(e)는 HMM의 스테이트열을 각각 나타낸다.

도 3은 본 발명에 따른 음성 인식 방법을 나타내는 플로우 챠트로서, 동일한 가중치를 갖는 특성 벡터에 문맥 의존적인 가중치를 프레임별로 부여하는 단계(제301~303단계) 및 학습 모델들에 근거하여 탐색하여 문맥 의존적인 가중치를 갖는 특성 벡터에 대한 음성을 인식하는 단계(제305단계)로 이루어진다.

이제, 도 1 내지 도 3을 참조하여, 본 발명에 따른 음성 인식 방법을 상세히 설명한다.

도 1 내지 도 3을 참조하여, 특성 벡터 추출부(20)는 도 2(a)에 도시된 화자의 음성 신호를 받아들이고, 고정 프레임 율로 음성 신호의 특성 벡터를 추출한 후, 각 프레임별로 동일한 가중치를 부여한다(제301단계). 즉, 특성 벡터 추출부(20)는 도 2(b)에 도시된 바와 같이 일정한 간격을 갖는 윈도우를 이용하여 고정된 프레임율로 특성 벡터를 추출할 수 있다. 또한, 각 프레임별로 동일한 가중치를 부여하므로, 도 2(c)에 도시된 바와 같이 각 프레임별 정보율은 동일하게 된다.

제301단계 후에, 프레임 정보율 추출부(30)는 특성 추출부(20)로부터 출력되는 동일한 가중치를 갖는 특성 벡터을 받아들이고, 다층 퍼셉트론의 연결 가중치를 학습한 학습 결과에 의해 문맥 의존적인 스테이트별 가중치를 갖는 특성 벡터를 추출한다(제303단계). 그 결과 프레임 정보율 추출부(30)는 도 2(c)에 도시된 바와 같이 각 프레임별 정보율은 스테이트별로 부여된 가중치에 따라 달라진다.

제303단계 후에, 음성 인식부(22)는 문맥 의존적인 스테이트별 가중치를 갖는 특성 벡터에 상응하여 도 2(e)에 도시된 바와 같은 HMM 스테이트열을 추출한다. 또한, 음성 인식부(22)는 HMM 스테이트열에 상응하는 음성 모델을 학습 모델들(24,26,28)로부터 탐색하고, 탐색된 결과를 음성 인식된 결과로서 출력한다(제305단계). 이처럼, 음향 문맥에 의존하는 스테이트별 가중치가 적용되었을 때, HMM 스테이트열 각각에서의 출력 확률 분포(Pr)는 다음 수학식 1과 같이 나타날 수 있다.

$$\log Pr(o_t, s) = \log Pr(o_t, s) + \sum_{p} w_p(o_t, s) \log Pr(o_{t+p}, s)$$

여기서, o_t는 특성 벡터를 나타내고, s는 스테이트를 나타내고, $Pr(o_t, s)$는 스테이트 s에서 t번째 프레임의 특성 벡터 o_t를 발생시킬 확률을 나타낸다. 또한, $w_p(o_t, s)$는 음향 문맥 $O_t=(o_{t-p}, o_{t-p+1}, \ldots, o_t, \ldots, o_{t+p})$에 의한 스테이트 s에서의 스테이트별 가중치를 나타낸다.

한편, HMM 스테이트열 각각에서의 출력 확률 분포(Pr)가 수학식 1과 같으면, 소정의 스테이트열 S를 통해 관측열 O가 발생될 확률의 로그 값은 다음 수학식 2와 같이 나타낼 수 있다.

$$\log Pr(O, S) = \log Pr(S) + \sum_{t} \log Pr(o_t, s_t)$$

여기서, O는 관측열을 나타내고, S는 히든 마코프 모델의 스테이트열을 나타내고, λ는 히든 마코프 모델의 파라미터를 나타내고, W는 문맥 의존적인 스테이트별 가중치들의 집합을 나타낸다. 또한, w_s는 스테이트별 가중치를 나타내고, O_t는 음향 문맥을 나타내고, o_t는 t번째 프레임에서의 특징 벡터를 나타내고, s_t는 스테이트를 나타낸다. 또한, $w_s(O_t)$는 음향 문맥 O_t에 의한 스테이트 s에서의 스테이트별 가중치를 나타낸다.

수학식 1 및 2를 참조하면, 음성 인식부(22)는 각 HMM 스테이트에서의 출력 확률 분포에 각 스테이트별 가중치가 w_s가 적용되므로, 학습 모델들 사이에 분별력을 향상시킬 수가 있다.

예컨대, 음성 인식에서 혼동되는 모델들을 서로 분별할 때 결정적인 역할을 하는 HMM 스테이트들이 있다. 즉, 본 발명에 따른 음성 인식의 경우, 스테이트별 가중치가 관측 확률의 계산에 결합되므로 서로 경쟁되는 모델들 사이에 더 나은 분별력이 부여될 수 있다. 또한, 본 발명에 따른 음성 인식의 경우, 다중 퍼셉트론에 의해 문맥 의존적인 스테이트별 가중치가 적용된다. 즉, 음향 문맥, 즉 현재 프레임의 주변 변이성에 따라 스테이트별 가중치가 달리 적용되므로 효과적이다.

한편, 음향 문맥에 의존하여 스테이트별로 가중치가 반영되기 위해서는 음향 문맥들을 패턴으로 분류하거나, 또는 음향 문맥들을 직접 입력으로 받아 스테이트별 가중치를 출력으로 낼 수 있는 구조의 분류기가 필요하다. 먼저, 음향 문맥들을 벡터 양자화기를 통해 정해진 전체 코드워드들 중의 하나로 결정할 수 있으며, 결정된 코드워드에 대한 스테이트별 가중치들이 인식에 적용될 수 있다. 그러나, 벡터 양자화기는 양자화 에러때문에 입력으로 받을 수 있는 차원의 수가 제한될 수 밖에 없으며, 적절한 거리 척도의 결정에도 어려움이 있다. 반면, 본 발명에서 이용하는 다층 퍼셉트론은 상당히 높은 차원의 벡터들을 입력으로 받을 수 있는 구조적인 특징과, 다층 퍼셉트론의 연결 가중치들이 분별적 방법에 의해 학습된다는 장점을 가지고 있다.

도 4는 다층 퍼셉트론을 이용한 본 발명에 따른 문맥 의존적인 스테이트별 가중치의 학습 장치를 나타내는 블록도이다. 본 발명에 따른 문맥 의존적인 스테이트별 가중치 학습 장치는 음성 인식부(46), 가중치 추출부(44) 및 에러 정정부(54)를 포함하여 구성된다. 설명의 편의를 위해, 도 4에는 음소 모델(52), 언어 모델(50), 발음 사전(48)등에 대한 학습 모델들을 함께 도시하였다.

도 4를 참조하여, 음성 인식부(46)는 현재 인식하고자 하는 학습 문장(42)을 받아들이고, 가중치 추출부(44)로부터 출력되는 스테이트별 가중치에 상응하여 학습 모델들(52,50,48)로부터 후보 가설들을 탐색한다. 여기서, 학습 문장(42)은 학습하고자 하는 학습 문장에 대한 특징 벡터(o)와 인접한 특징 벡터들간의 차(Δo)에 대한 정보를 포함한다.

에러 정정부(54)는 올바른 가설에 대한 로그 확률 값과 음성 인식부(46)에서 탐색된 후보 가설에 대한 로그 확률값 사이의 거리(di)를 측정하여 오류 함수값(ei)을 계산한다.

다층 퍼셉트론(40)은 입력층, 출력층 및 중간층으로 구성되며, 현재 인식하고자 하는 학습 문장의 프레임을 중심으로 전/후 p개의 프레임을 관측열(O_t)로서 입력층으로 받아들인다. 또한, 오류 함수값(ei)에 상응하여 입력층, 출력층 및 중간층간의 연결 가중치를 조정한다. 여기서, 다층 퍼셉트론(40)의 출력층에서는 전체 모델들의 총 스테이트의 수만큼 노드(node)수를 가진다. 그리고, 출력층의 각 노드에서의 출력값들은 해당 노드가 가르키는 스테이트의 가중치를 나타낸다. 또한, 중간층의 노드수는 입력층의 노드수와 출력층의 노드수 그리고, 다층 퍼셉트론의 학습 가능성을 고려하여 적절히 결정할 수 있다. 또한, 다층 퍼셉트론(40)은 가중치 추출부(44)에서 출력되는 스테이트별 가중치들이 초기적으로 1에 가까운 값으로 초기화시키기 위해, 연결 가중치들을 0에 가까운 작은 값으로 초기화된다.

가중치 추출부(44)는 다층 퍼셉트론(40)의 연결 가중치에 상응하여 현재 인식하고자 하는 학습 문장의 스테이트별 가중치를 추정하고, 추정된 가중치를 음성 인식부(46)로 출력한다.

도 5는 다층 퍼셉트론을 이용한 본 발명에 따른 문맥 의존적인 스테이트별 가중치 학습 방법을 나타내는 플로우 챠트로서, 문맥 의존적인 스테이트별 가중치들을 초기화한 후, 다수개의 학습 문장들에 대한 후보 가설들을 탐색하는 단계(제501~503단계), 오류 함수값을 구하고, 오류 함수값에 따라 다층 퍼셉트론의 가중치를 조정하는 단계(제505~507단계) 및 모든 학습 문장에 대해 수렴조건이 만족될 때까지 제501~507단계를 반복하는 단계(제509~511단계)로 이루어진다.

도 4 및 도 5를 참조하면, 문맥 의존적인 스테이트별 가중치들이 1에 가까운 값으로 초기화되도록 다층 퍼셉트론(40)의 연결 가중치들을 0에 가까운 값으로 초기화한다(제501단계). 제501단계 후에, 학습 모델들(52,50,48)로부터 다수개의 학습 문장들에 대한 후보 가설들을 탐색한다(제503단계).

제503단계 후에, 에러 정정부(54)는 올바른 가설에 대한 로그 확률값과 후보 가설에 대한 로그 확률값 사이의 거리(d_i)를 측정하여 에러 함수값(e_i)을 구한다(제505단계). 이 때, 제505단계에서 올바른 가설에 대한 로그 확률값과 후보 가설에 대한 로그 확률값 사이의 거리(d_i)는 다음 수학식 3에 의해 구해질 수 있다.

$$d_i = \log Pr(O_t \mid \lambda_{correct}, C) - \log Pr(O_t \mid \lambda_{candidate}, C)$$

여기서, $\log Pr(O_t \mid \lambda_{correct}, C)$는 올바른 가설에 대한 로그 확률값을 나타내고, $\log Pr(O_t \mid \lambda_{candidate}, C)$는 후보 가설에 대한 로그 확률값을 나타내며, O_t는 다층 퍼셉트론(40)으로 입력되는 p개의 관측열을 나타낸다. 또한, $\lambda_{correct}$는 올바른 가설에 대한 모델을 나타내고, $\lambda_{candidate}$는 후보 가설에 대한 모델을 나타내고, C는 다층 퍼셉트론(40)의 연결 가중치들의 집합을 나타낸다.

또한, 제505단계에서 오류 함수값(e_i)은 비선형 함수인 시그모이드 함수가 이용되며, 다음 수학식 4에 의해 구해질 수 있다.

$$\frac{1}{1+\exp(-\alpha x)}$$

여기서, α는 시그모이드 함수의 기울기를 나타낸다.

계속해서, 제505단계에서 오류 함수값(e_i)가 구해지면 오류 함수값(e_i)에 상응하여 다층 퍼셉트론의 연결 가중치를 조정한다(제507단계). 이 때, 제507단계는 다음 수학식 5에 의해 다층 퍼셉트론(40)의 연결 가중치를 조정한다.

$$C_k = C_{k-1} + \varepsilon_k \nabla E(C_{k-1})$$

여기서, C_k 및 C_{k-1}는 각각 k번째 및 k-1번째 반복 횟수에서 추정된 다층 퍼셉트론의 연결 가중치들의 집합을 나타내고, ε_k는 k번째 반복 횟수에서의 학습률을 나타낸다. 또한, $\nabla E(C_{k-1})$는 k번째 반복 횟수에서 전체 학습 문장들에 대한 오류를 나타내고, λ는 제503단계에서 학습 데이터 베이스로부터 다수개의 학습 문장들에 대한 후보 가설들을 탐색하는 음성 인식기의 파라미터를 각각 나타낸다.

계속해서, 제507단계 후에, 다수개의 학습 문장들에 대한 다층 퍼셉트론(40)의 연결 가중치가 모두 조정되었는가를 판단한다(제509단계). 제509단계에서, 다수개의 학습 문장들에 대한 다층 퍼셉트론(40)의 연결 가중치가 모두 조정되었다고 판단되면, 에러 함수값(e_i)이 소정값 이하인가를 판단하고, 에러 함수값(e_i)이 소정값 이하로 수렴될 때까지 제501~509단계를 반복 수행한다(제511단계). 결국, 다층 퍼셉트론(40)의 연결 가중치들은 오류 함수값(e_i)이 최소화되는 방향으로 반복적인 과정을 통해 학습된다.

한편, 제511단계는 에러 함수값(e_i)의 수렴여부 대신, 제503단계 내지 제509단계가 소정 횟수만큼 반복 진행되었는가를 판단할 수도 있다. 즉, 제511단계는 소정의 반복 횟수만큼 제503단계 내지 제509단계가 반복 진행되면, 다층 퍼셉트론의 연결 가중치 학습을 종료할 수 있다.

이상에서와 같이, 본 발명은 다층 퍼셉트론을 이용하여 문맥 의존적인 스테이트별 가중치를 학습하고, 학습된 결과를 음성 인식에 이용한다. 따라서, 기존의 음성 인식기에서 상대적으로 길고 안정된 음성 구간들에 의해 결과가 결정되는 단점을 보완할 수 있다. 또한, 기존의 음성 인식기에서 상대적으로 길고 안정된 음성 구간들에 의해 모델들이 학습되고 인식되는 문제를 보상할 수 있으며, 기존의 음성 인식기에서 각 음성 구간의 정보적 중요성을 반영하지 못하는 문제점을 보완할 수 있다.

이상에서 기술된 본 발명에 따른 음성 인식 장치 및 방법의 성능을 평가하기 위하여 두가지 종류의 음성 데이터 베이스에 적용하는 실험을 하였다. 실험에 사용된 음성 데이터 베이스로는 음소가 균형을 이룬 445개 어휘의 고립단어 태스크와 생활 정보 안내와 관련된 어휘수가 약 1100개인 연속음성 태스크이다. 실험에서 사용된 특징 벡터는 12차의 PLP(perceptually linear prediction) 계수와 에너지, 그리고 이들의 차분 계수로서 한 프레임이 26차의 벡터로 표현된다. 음성 신호는 음소 모델을 사용하여 모델링 되었으며, 각 음소 모델은 3개의 상태를 가지는 간단한 좌우향 모델 구조를 가지고 있다. 단, 단어 사이의 묵음과 문장의 처음과 끝에서의 묵음 모델은 스테이트의 수가 1개이다. 각 상태에서의 확률 분포는 16개의 혼합 성분을 사용한 혼합 가우시안 밀도 함수에 의해 표현되었다.

다층 퍼셉트론의 입력층에서는 현재 프레임과 좌, 우 2프레임씩, 총 130차원의 벡터가 사용된다. 다층 퍼셉트론의 연결 가중치들은 각 학습 문장마다 온라인으로 조정되었다. 오류 함수값을 올바른 가설의 로그 확률값과 오인식된 가설의 로그 확률값의 차이에 대한 비선형 함수로 정의 함으로써, 학습 데이터 집합에 대한 오류의 수가 최소가 되도록 다층 퍼셉트론의 노드간 연결 가중치들을 조정한다.

도 6(a)~(d)는 본 발명에서와 같이 문맥 의존적인 스테이트별 가중치가 적용된 경우와 적용되지 않은 경우의 음성 인식 결과를 나타내는 도면으로, 도 6(a)는 인식하고자 하는 음성 신호를 나타내고, 도 6(b)는 종래의 음성 인식기에서 올바른 가설과 오인식된 가설간의 프레임별 로그 확률값을 나타내고, 도 6(c)는 올바른 가설과 오인식된 가설의 스테이트열에 대한 문맥 의존적인 스테이트별 가중치의 변화를 보이고, 도 6(d)는 문맥 의존적인 스테이트별 가중치가 적용된 후의 올바른 가설과 오인식된 가설간의 프레임별 로그 확률값의 차이를 나타낸다.

참고로, 실험에 사용된 입력 음성은 '이십일일 스무시'라는 발성 음성이다. 그러나, 종래와 같이 문맥 의존적인 스테이트별 가중치를 적용하지 않은 경우 '이십이일 스무시'라고 오인식되었다. 즉, 음소 /OI/와 /??/구간을 제대로 구별하지 못한 결과이다. 도 6(b)를 참조하면, 종래의 음성 인식기는 오인식된 부분(A)에서 올바른 가설(105)과 오인식된 가설(100)간에 다른 로그 확률값을 보임을 알 수 있다. 즉, 도 6(c)를 참조하면, 올바른 가설('이십일일 스무시') 및 오인식된 가설('이십이일 스무시')간의 문맥 의존적인 스테이트별 가중치(115 및 110)가 오인식된 부분(A)에서 서로 상반됨을 보인다. 결국, 문맥 의존적인 스테이트별 가중치를 적용할 경우, 오인식되는 오류를 수정할 수 있으며, 따라서 전체적인 음성 인식의 성능을 개선할 수 있다.

도 7은 고립단어 태스크와 연속언어 태스크에 대해 종래 인식기와 본 발명간의 음성 인식 결과를 나타내는 도면이다.

도 7을 참조하면, 고립 단어 대스크에 대한 음성 인식 테스트 결과, 가중치를 적용하지 않은 종래의 인식기(200)는 9.9%의 단어 오인식률을 나타낸다. 또한, 종래의 음향 문맥에 독립적인 고정된 스테이트별

가중치가 적용된 인식기(210)의 경우에는 9%의 다어 오인식률을 나타낸다. 반면, 본 발명에서와 같이 문맥 의존적인 스테이트별 가중치를 적용한 인식기(220)의 경우에는 6.8%의 단어 오인식률을 얻을 수 있다. 즉, 본 발명에서와 같이 문맥 의존적인 스테이트별 가중치를 적용한 경우 단어 오인식률이 개선됨을 알 수 있다.

또한, 연속어 태스크에 대한 음성 인식 테스트에서도 가중치를 적용하지 않거나 또는 고정된 가중치를 적용한 종래의 인식기들(230,240)에 비해, 본 발명의 문맥 의존적인 스테이트별 가중치를 적용한 인식기(250)의 경우 오인식률이 낮음을 보인다.

결국, 주변 음성의 변이 정도에 따라 스테이트별 가중치를 달리 부여하는, 즉, 정보적 중요도를 반영하는 음성 인식 방법이 음성 인식 시스템의 성능을 크게 개선할 수 있다.

발명의 효과

상술한 바와 같이 본 발명에 따른 음성 인식 장치 및 방법은 주변 음성의 변이 정도에 따라 스테이트별 가중치를 달리 부여하므로, 스테이트별로 정보적 중요도를 반영하여 음성을 인식하므로 음성의 오인식률을 크게 줄일 수 있다.

(57) 청구의 범위

청구항 1

음향 문맥에 상응하여 특성 벡터마다 히든 마코프 스테이트별로 특성 벡터의 중요성을 판단하여 화자의 음성을 인식하는 음성 인식 장치에 있어서,

상기 화자의 음성 신호를 받아들이고, 고정 프레임 율로 상기 음성 신호의 특성 벡터를 추출하고, 추출된 프레임별 특성 벡터에 동일한 가중치를 부여하는 특성 추출부;

학습 데이터 베이스로부터 분별적 방법에 의해 학습된 다층 퍼셉트론에 의해 상기 프레임마다 히든 마코프 스테이트별 가중치를 적용하여, 상기 프레임별로 문맥 의존적인 정보율을 갖는 특성 벡터를 발생하는 프레임 정보율 추출부; 및

히든 마코프 모델과 학습 모델들을 이용하여 상기 학습된 다층 퍼셉트론으로부터 추정된 문맥 의존적인 스테이트별 가중치를 갖는 특성 벡터를 인식하는 음성 인식부를 포함하는 것을 특징으로 하는 스테이트별 가중치를 적용한 음성 인식 장치.

청구항 2

음향 문맥에 따라 히든 마코프 스테이트별로 특성 벡터의 중요도를 판단하여 화자의 음성을 인식하는 음성 인식 방법에 있어서,

(a)상기 화자의 음성 신호로부터 고정 프레임 율로 특성 벡터를 추출하고, 추출된 프레임별 특성 벡터에 동일한 가중치를 부여하여 동일한 스테이트별 가중치를 갖는 프레임별 특성 벡터를 생성하는 단계;

(b)학습 데이터 베이스에 대해 분별적 방법에 의해 학습된 다층 퍼셉트론에 의해 상기 (a)단계에서 추출된 프레임마다 히든 마코프 스테이트별 가중치를 부여하여 문맥 의존적인 스테이트별 가중치를 갖는 특성 벡터를 생성하는 단계; 및

(c)히든 마코프 모델과 학습 모델들을 이용하여 상기 학습된 다층 퍼셉트론으로부터 추정된 문맥 의존적인 스테이트별 가중치를 갖는 특성 벡터를 인식하는 단계로 이루어지는 것을 특징으로 하는 스테이트별 가중치를 적용한 음성 인식 방법.

청구항 3

제2항에 있어서, 상기 (c)단계의 상기 히든 마코프 모델은 다음 수학식 1에 나타난 확률 분포값(Pr)을 갖고,

[수학식 1]

$$\log \Pr(O, S \mid \lambda) = \log \Pr(S \mid \lambda) + \sum_{t} w_{s_t}(O_t) \log o_{t,s_t}$$

여기서, O는 관측열을 나타내고, S는 상기 히든 마코프 모델의 스테이트열을 나타내고, λ는 상기 히든 마코프 모델의 파라미터를 나타내고, W는 문맥 의존적인 스테이트별 가중치들의 집합을 나타내고, w_s는 스테이트별 가중치를 나타내고, O_t는 음향 문맥을 나타내고, o_t는 t번째 프레임에서의 특징 벡터를 나타내고, s_t는 스테이트를 나타내고, $w_s(O_t)$는 음향 문맥 O_t에 의한 스테이트 s에서의 스테이트별 가중치를 각각 나타내는 것을 특징으로 하는 스테이트별 가중치를 적용한 음성 인식 방법.

청구항 4

학습을 통해 문맥 의존적인 스테이트별 가중치를 추정하는 가중치 학습 장치에 있어서,

현재 인식하고자 하는 학습 문장에 대한 특성 벡터와 스테이트별 가중치에 상응하여 복수개의 후보 가설들을 탐색하는 음성 인식부;

올바른 가설에 대한 확률 값과 상기 음성 인식부에서 탐색된 후보 가설들에 대한 확률값 사이의 거리를 측정하여 오류 함수값을 계산하는 에러 정정부;

입력층, 출력층 및 중간층으로 구성되며, 현재 인식하고자 하는 학습 문장의 프레임을 중심으로 전/후 p개의 프레임을 받아들이고, 상기 오류 함수값에 상응하여 상기 입력층, 상기 출력층 및 상기 중간층간의 연결 가중치들을 조정하는 다층 퍼셉트론; 및

상기 다층 퍼셉트론의 연결 가중치에 상응하여 현재 인식하고자 하는 학습 문장의 상기 스테이트별 가중치를 추정하고, 추정된 상기 스테이트별 가중치를 상기 음성 인식부로 출력하는 가중치 추정부를 포함하는 것을 특징으로 하는 문맥 의존적인 스테이트별 가중치 학습 장치.

청구항 5

다층 퍼셉트론을 이용하여 학습 문장의 문맥 의존적인 스테이트별 가중치를 추정하는 가중치 학습 방법에 있어서,

(a)문맥 의존적인 스테이트별 가중치들이 1에 가까운 값으로 초기화되도록 상기 다층 퍼셉트론의 연결 가중치들을 0에 가까운 값으로 초기화하는 단계;

(b)현재 인식하고자 하는 학습 문장에 대한 특성 벡터와 스테이트별 가중치에 상응하여 학습 데이터 베이스로부터 복수개의 후보 가설들을 탐색하는 단계;

(c)올바른 가설에 대한 확률값과 상기 (b)단계에서 탐색된 후보 가설들에 대한 확률값 사이의 거리를 측정하여 에러 함수값을 구하는 단계;

(d)상기 에러 함수값에 상응하여 상기 다층 퍼셉트론의 연결 가중치를 조정하는 단계;

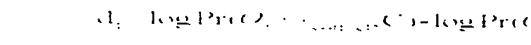
(e)다수개의 학습 문장들에 대한 상기 다층 퍼셉트론의 연결 가중치가 모두 조정되었는가를 판단하는 단계; 및

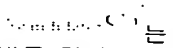
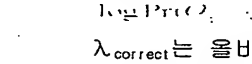
(f)상기 (e)단계에서 상기 다수개의 학습 문장들에 대한 상기 다층 퍼셉트론의 연결 가중치가 모두 조정되었다고 판단되면, 에러 함수값이 소정값 이하로 될 때까지 상기 (b) 내지 (e)단계를 반복 진행하는 단계로 이루어지는 것을 특징으로 하는 다층 퍼셉트론을 이용한 문맥 의존적인 스테이트별 가중치 학습 방법.

청구항 6

제5항에 있어서, 상기 (c)단계는 다음 수학식 1에 의해 올바른 가설에 대한 확률값과 후보 가설에 대한 확률값 사이의 거리(d_i)를 측정하고,

[수학식 1]

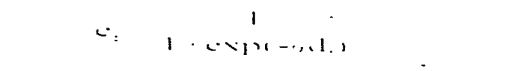


여기서,  는 올바른 가설에 대한 로그 확률값을 나타내고,  는 후보 가설에 대한 로그 확률값을 나타내며, O_i는 i번째 관측열을 나타내고, $\lambda_{correct}$는 올바른 가설에 대한 모델을 나타내고, $\lambda_{candidate}$는 후보 가설에 대한 모델을 나타내고, C는 다층 퍼셉트론의 연결 가중치들의 집합을 나타내는 것을 특징으로 하는 다층 퍼셉트론을 이용한 문맥 의존적인 스테이트별 가중치 학습 방법.

청구항 7

제5항에 있어서, 상기 (c)단계의 상기 오류 함수값은 비선형 함수인 시그모이드 함수가 이용되며, 다음 수학식 2에 의해 상기 에러 함수값(e_i)을 구하고,

[수학식2]

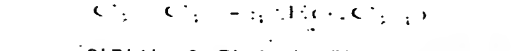


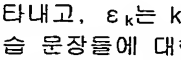
여기서, α는 상기 시그모이드 함수의 기울기를 나타내는 것을 특징으로 하는 다층 퍼셉트론을 이용한 문맥 의존적인 스테이트별 가중치 학습 방법.

청구항 8

제5항에 있어서, 상기 (d)단계는 다음 수학식3에 의해 상기 다층 퍼셉트론의 연결 가중치를 조정하고,

[수학식 3]



여기서, C_k 및 C_{k-1}는 각각 k번째 및 k-1번째 반복 횟수에서 추정된 상기 다층 퍼셉트론의 연결 가중치들의 집합을 나타내고, ε_k는 k번째 반복 횟수에서의 학습률을 나타내며,  는 k번째 반복 횟수에서 전체 학습 문장들에 대한 오류를 나타내고, λ는 상기 (b)단계에서 학습 데이터 베이스로부터 상기 다수개의 학습 문장들에 대한 후보 가설들을 탐색하는 음성 인식기의 파라미터를 각각 나타내는 것을 특징으로 하는 다층 퍼셉트론을 이용한 문맥 의존적인 스테이트별 가중치 학습 방법.

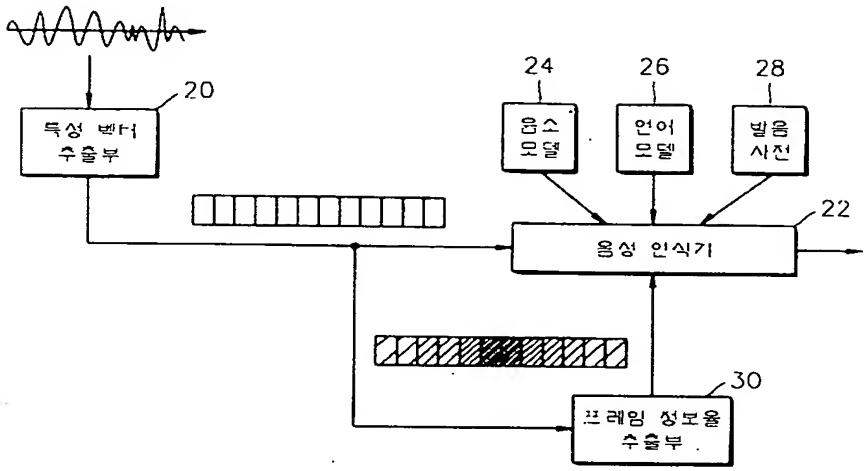
청구항 9

제5항에 있어서, 상기 (f)단계는 상기 다수개의 학습 문장들에 대한 상기 다층 퍼셉트론의 연결 가중치

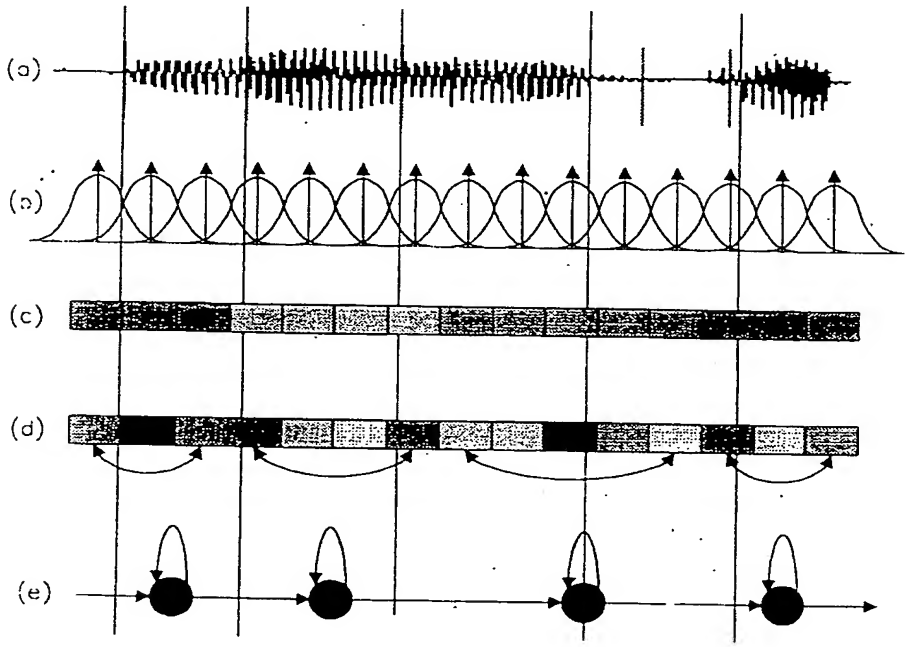
가 모두 조정되었다고 판단되면, 소정의 반복 횟수만큼 상기 (b)단계 내지 상기 (e)단계가 반복 진행하는 단계로 대체될 수 있는 것을 특징으로 하는 다층 퍼셉트론을 이용한 문맥 의존적인 스테이트별 가중치 학습 방법.

도면

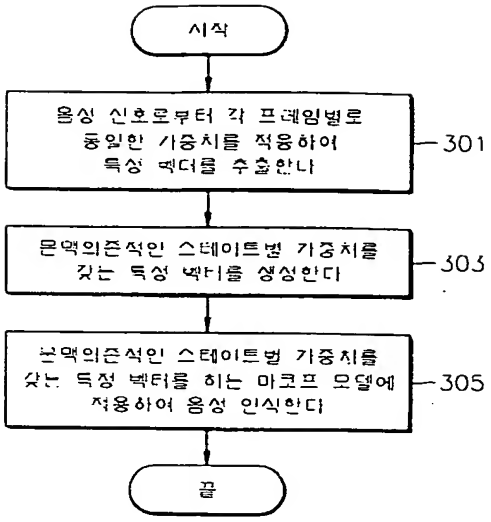
도면1



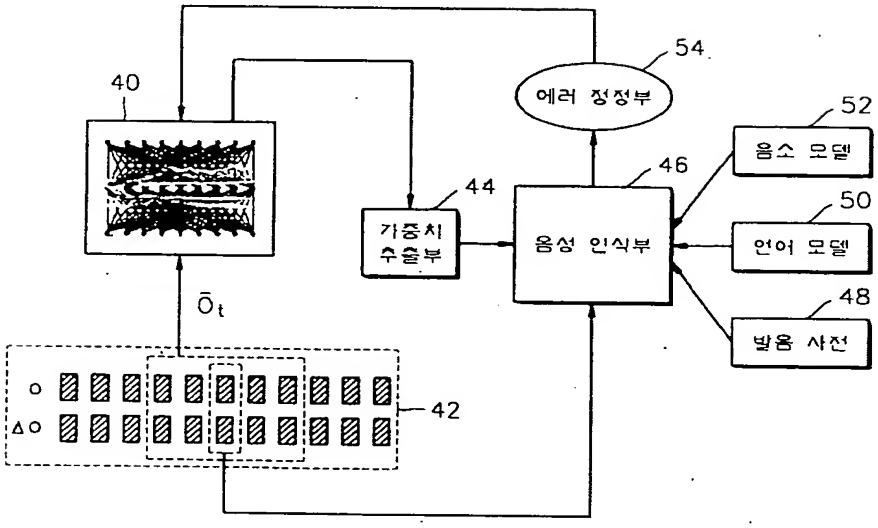
도면2



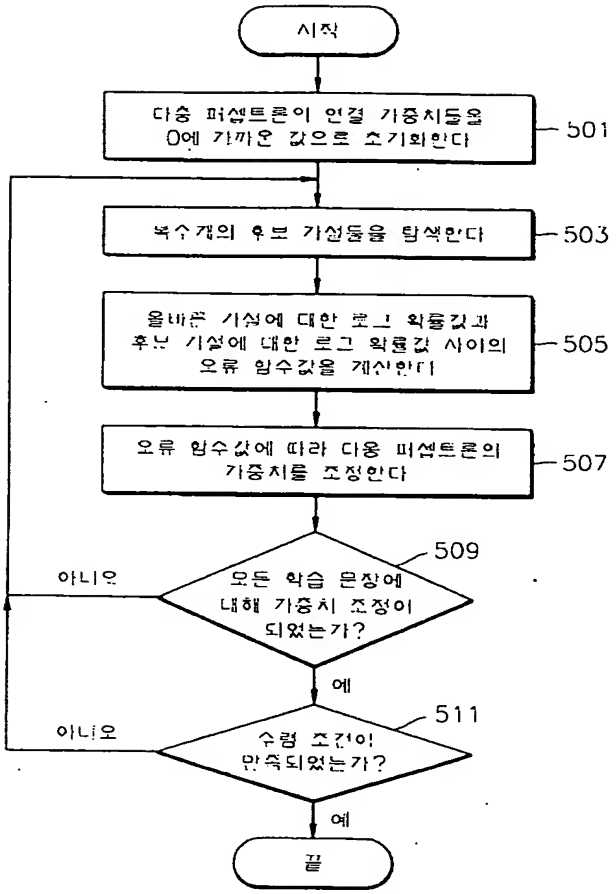
도면3


시작
음성 신호로부터 각 프레임별로 동일한 가중치를 적용하여 특성 벡터를 추출한다
301
문맥의존적인 스테이트별 가중치를 갖는 특성 벡터를 생성한다
303
문맥의존적인 스테이트별 가중치를 갖는 특성 벡터를 히든 마코프 모델에 적용하여 음성 인식한다
305
끝

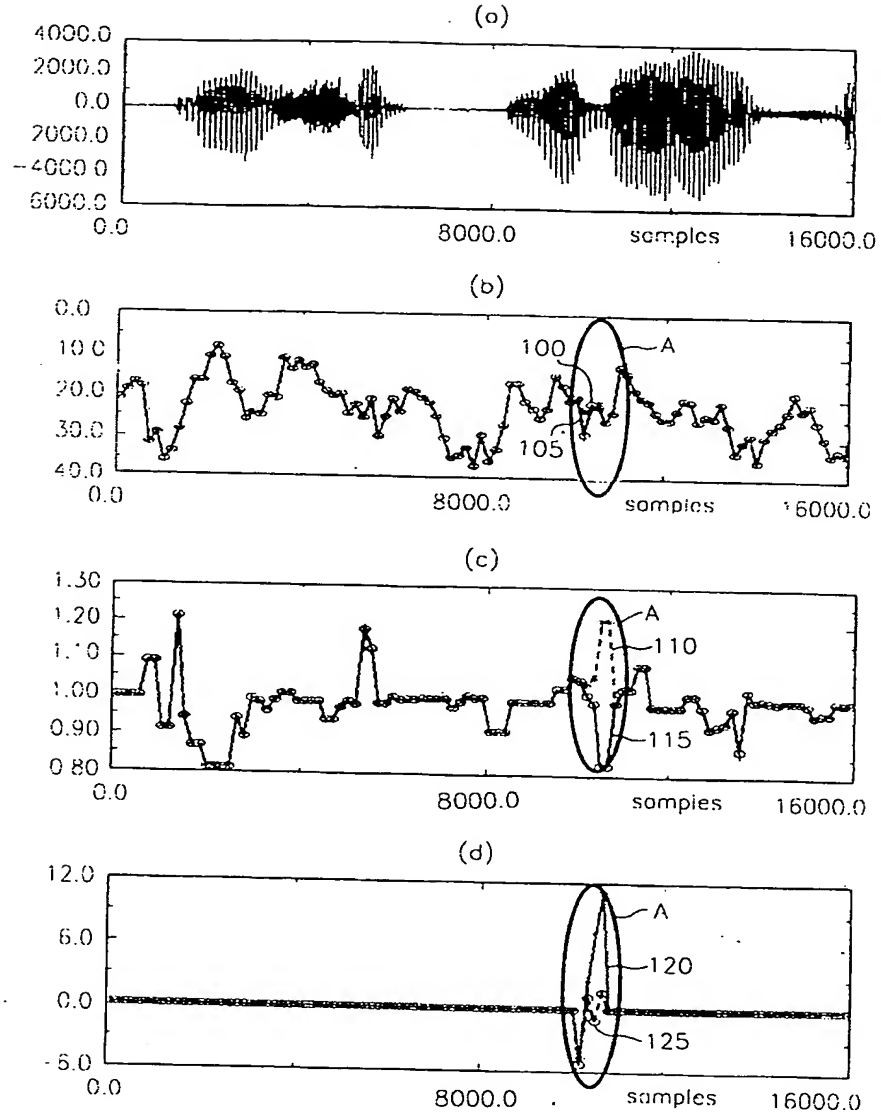
도면4


40
54
에러 정정부
52
음소 모델
46
44
가중치 추출부
음성 인식부
50
언어 모델
48
발음 사전
$\bar{O}_t$
42

도면5


시작
다중 퍼셉트론의 연결 가중치들을 0에 가까운 값으로 초기화한다
501
복수개의 후보 가설들을 탐색한다
503
올바른 가설에 대한 로그 확률값과 후보 가설에 대한 로그 확률값 사이의 오류 함수값을 계산한다
505
오류 함수값에 따라 다중 퍼셉트론의 가중치를 조정한다
507
509
모든 학습 문장에 대해 가중치 조정이 되었는가?
아니오
예
511
수렴 조건이 만족되었는가?
아니오
예
끝

도면6


(a)
4000.0
2000.0
0.0
-2000.0
-4000.0
-6000.0
0.0
8000.0
samples
16000.0
(b)
0.0
10.0
20.0
30.0
40.0
0.0
8000.0
samples
16000.0
100
A
105
(c)
1.30
1.20
1.10
1.00
0.90
0.80
0.0
8000.0
samples
16000.0
A
110
115
(d)
12.0
6.0
0.0
-6.0
0.0
8000.0
samples
16000.0
A
120
125

도면7